PROCESS KNOWLEDGE AND DATA QUALITY OUTCOMES

Yang W. Lee
Northeastern University
y.lee@neu.edu

Diane M. Strong
Worcester Polytechnic Institute
dstrong@wpi.edu

Abstract What modes and domains of knowledge about data production processes are most critical for producing high-quality data? This study provides an answer to this question. Data are collected via questionnaire and analyzed using linear regression. The results show some similarities and differences in which knowledge variables are significant for various data quality dimensions. Three results are of particular interest to data quality managers and researchers. The first is the complexity and mix of knowledge associated with producing accurate data. The second is the significant results overall for knowledge about the data collection process as compared to data storage and utilization processes. The third is the negative associations of the knowing-why mode of knowledge as compared to the positive associations for knowing-what and knowing-how. Each of these results has managerial implications and generates avenues for further research.

1. INTRODUCTION

Data quality research is moving beyond understanding the data quality construct itself to studies that focus on understanding what contributes to high quality data. In so doing, researchers are investigating the process of producing data to help understand the variables that contribute to high and low quality data. For example, recent studies have viewed information as a product produced by an information (or data[1]) production process [12] [1]. This approach focuses on the process of producing data and information, rather than only on the quality of the information product produced.

In this study, we examine workers' knowledge about the data production process as a contributor to the quality of data produced by the process. We are taking a process view of data quality. Data, or information products, are produced by a process that starts with the collection of raw data and ends with the utilization of information products by information consumers working on various tasks. The quality of the data or information product produced is determined by the activities performed as part of the data production process. According to the conventional wisdom about processes, the ability to meet the goals and objectives of a process depends on workers who are knowledgeable about the entire process, beyond their individual work activities [2].

---

[1] While we occasionally mention raw data inputs and final information products for consumers, most of this paper refers to the process by which data is converted into information. When referring to this process or portions of it, the terms "data" and "information" are used interchangeably.

Our basic conjecture is that more knowledgeable workers will contribute to higher quality data. Beyond this general conjecture, we investigate which modes and domains of knowledge contribute to which dimensions of data quality. We also guard against "intellectual monism" [9], which might stress only one side of the general conjecture. In Schultz and Leider's recent review of IS researchers' work [9] in knowledge management, they interpret March [8] and others as follows [9]:
"…Knowledge is a double-edged sword: too little leads to inefficiencies, too much results in rigidities that tend to be counterproductive in a dynamically changing world: … too little might result in …. expensive mistakes, and too much might result in unwanted accountability."

This study is a step towards explaining the relationship between process knowledge and data quality outcomes. This is also a first step towards determining training needs for workers performing activities in a data production process in practice.

2. RESEARCH MODEL

For our investigation of the relationship between the quality of data and the knowledge of workers performing activities to produce that data, we start with a simple model involving two sets of variables, those capturing the dimensions of data quality and those capturing the knowledge domains and modes of process workers.

2.1 Data Quality Dimensions

High data quality means data that are fit for use by data consumers [13] [10]. In this study, we focus on five dimensions of data quality: Accuracy, Relevancy, Timeliness, Completeness, and Accessibility. ***Accuracy*** denotes the extent to which data are correct and free-of-error. ***Relevancy*** denotes the extent to which data are applicable and useful for the task at hand. ***Timeliness*** denotes the extent to which the data are sufficiently up-to-date for the task at hand. ***Completeness*** denotes the extent to which data are not missing and are of sufficient breadth and depth for the task at hand. ***Accessibility*** denotes the extent to which data are available, or easily and quickly retrievable. Each of these quality dimensions is a performance goal of the data production process. That is, the overall goal is to produce accurate, complete, and timely data that are accessible to data consumers and relevant to their tasks.
These five dimensions are a subset of the data quality dimensions found to be important to data consumers [13]. We chose these five dimensions for our study rather than the full sixteen in [4] for several reasons. First, these five provide a reasonable sample of data quality dimensions since they represent at least one dimension from each of the four higher level categories: Soundness includes Accuracy, Usefulness includes Relevance and Completeness, Dependability includes Timeliness, and Usability includes Accessibility [4]. Second, these five dimensions are generally of critical importance to users. Third, while other dimensions, such as consistent representation and understandability, are clearly important dimensions of data quality, testing all sixteen is not necessary for investigating the role of knowledge in producing high quality data, and may only complicate the presentation.

2.2 Data Quality Process Knowledge

Three modes of knowledge are pertinent for producing process outcomes that meet objectives, knowing-what, knowing-how, and knowing-why [5] [6]. In the context of a data production process, **knowing-what** is defined as understanding the activities involved in the process. **Knowing-how** is defined as understanding procedures to handle known data quality problems. **Knowing-why** is defined as the ability to analyze underlying principles and discover previously unknown data quality problems and solutions [5].

Like the production of physical products, a data production process, too, is divided into distinctive work processes. Three such data production processes are collection, storage, and utilization work-processes [10] [3]. The three modes of knowledge apply to these three processes. Thus, we can talk about knowing-what about collection, knowing-what about storage, and knowing-what about utilization, and similarly for knowing-how and knowing-why. As a result, there are nine relevant knowledge variables. See Lee and Strong [5] for a theoretical development of these nine knowledge variables.

2.2 Relationship of Knowledge to Data Quality

To investigate our conjecture that workers' knowledge about data production processes contributes to the quality of data produced from these processes, we collect data to estimate the following five relationships:

$DQ_i = a_i + \Sigma_{m,p}\, b_{imp}K_{mp} + e_i$

Where DQ is Data Quality
K is Knowledge
i=1..5, Data Quality Dimension (Accuracy, Relevancy, Timeliness, Completeness, Accessibility)
m=1..3, Mode of Knowledge (knowing-what, -how, and -why)
p=1..3, Data Production Process (collection, storage, utilization)

That is, we are investigating a simple linear model of how various modes and process domains of knowledge affect each dimension of data quality. Each dimension of data quality is independently examined. While this is a simple model, linear models are a good starting point for investigating relationships, and will provide input for further analysis as needed.

3. METHOD

3.1 Sites and Sample

Six companies served as data collection sites for this research, two financial institutions, three health care organizations, and a consumer product manufacturing company. In all six companies the research focus was the quality of their customer activity data. Specifically, in the financial institutions, the focus was the quality and production of data about investors and their investment activity. In the healthcare organizations, it was the quality and production of data about patients and the healthcare services they used. In the manufacturing firm, it was customers and their purchasing activity.

The sample consisted of 155 respondents from these six companies. Respondents were selected to ensure coverage of all the roles in the data production process (data collectors, data custodians, and data consumers) for the customer activity data being studied. The sample consisted of 48 data collectors, 45 data custodians, and 62 data consumers.

3.2 Data Collection Instrument and Procedures

The individuals in these companies selected for potential participation were invited to a one-hour session. First, an overview of the research and the questionnaire was presented. This overview motivates subjects to complete the questionnaire carefully and thoroughly. Pilot tests of the questionnaire indicated that such a session was necessary to ensure that the questionnaire was completed in full with quality responses. Second, the respondents were instructed to fill out the questionnaire for the customer activity database

and its associated data production process. Finally, after this presentation, respondents completed the questionnaire before leaving the session, ensuring nearly a 100% response rate.
Knowledge, the independent variable, is measured for each of the three modes, knowing-what, knowing-how, and knowing-why, applied to each of the three data production processes, data collection, data storage, and data utilization. Thus, there are nine sub-measures of knowledge. Data quality, the dependent variable, is measured for five key dimensions of importance to data consumers, accuracy, completeness, timeliness, relevance, and accessibility.

All independent and dependent variables were collected by questionnaire. Specifically, respondents assessed data quality along five dimensions and assessed their knowledge along nine knowledge constructs. The data quality measures have been well tested in previous research [11] [3], as were the knowledge measures [3] [5]. Four to seven items were included for each construct (See Appendix B for the items for each construct). For the nine knowledge measures, the scale was 1 to 10, where 1 is "Very small extent", 5 is "Average", and 10 is "Very Large Extent." For the five data quality measures, the scale is 0 to 10, where 0 is "Not at all", 5 is "Average", and 10 is "Completely." The measure for each construct is the mean value of the associated item measures. Cronbach alphas for the constructs ranged from .87 to .96, all showing acceptable reliability with no problem constructs.

3.3 Analysis

The models were fit using linear regression in SAS. First, all variables were tested for normality, and no problems were found. Second, the five overall models with all nine knowledge variables were estimated. All models were significant with random variance, but not all of the knowledge variables were significant. Third, to determine which variables were significant, a variable selection method, specifically backward elimination, was used to find the best model with significant independent variables. As a result, all five models were significant, with all variables significant at least at 0.1 (the elimination criterion). The residual plots indicated randomness and the QQ-plot shows no divergence from normality. These final models are the ones presented and discussed in this paper. Finally, for these final models, we checked for the significance of various demographic variables including the role of the respondent (data collector, data custodian, data consumer) and for the significance of the company indicator.

4. FINDINGS

Table 1 displays the key findings from the five regression models in a way that shows the patterns in significant knowledge variables across the five regressions. The full statistical results from the five regressions are in Appendix A. For the accuracy dimension, five of the nine knowledge variables are significant. For the other four data quality dimensions, only three knowledge variables are significant. These results indicate that process knowledge is clearly important for producing high-quality data. Beyond this, there are similarities and differences in which knowledge dimensions are significant for a particular data quality dimension. We further discuss three patterns in these results: (1) the similarities and differences by DQ dimension, (2) the similarity in the importance of knowledge about data collection, and (3) the inverse relationship for knowing-why.

4.1 Patterns by Data Quality Dimension

Of the five data quality dimensions investigated, the relationship between knowledge and accuracy is clearly the most complex. Producing data of high accuracy requires knowledge about the collection process, the storage process, and the utilization process. It also requires all three modes of knowledge, knowing-what, knowing-how, and knowing-why.

The relevancy dimension is the only other dimension for which knowledge of the utilization process is significant. This follows logically from the definition of relevancy. To produce data that are highly relevant, knowledge is needed about what activities and data are used during the utilization process. Data storage knowledge is not significant for relevancy, nor is knowing-how to find problems during the process. Both of these results are consistent with the meaning of the relevance dimension.

Accuracy	**Collection process**	**Storage process**	**Utilization process**
Knowing-what		0.408 (0.123)	
Knowing-how	0.408 (0.180)		0.192 (0.107)
Knowing-why	-0.546 (0.179)	-0.240 (0.132)	
Relevancy			
Knowing-what	0.413 (0.125)		0.280 (0.094)
Knowing-how			
Knowing-why	-0.198 (0.112)		
Timeliness			
Knowing-what		0.244 (0.083)	
Knowing-how	0.477 (0.169)		
Knowing-why	-0.469 (0.174)		
Completeness			
Knowing-what		0.266 (0.073)	
Knowing-how	0.496 (0.149)		
Knowing-why	-0.475 (0.154)		
Accessibility			
Knowing-what		0.214 (0.085)	
Knowing-how	0.596 (0.175)		
Knowing-why	-0.451 (0.180)		

Table 1: Regression Findings for Data Quality and Knowledge

An interesting pattern in knowledge significance by dimension is the similar pattern across the timeliness, completeness, and accessibility dimensions. All three have the same three significant knowledge variables, knowing-how about data collection, knowing-why about data collection, and knowing-what about storage.

4.2 Patterns by Knowledge Domain

Considering the domain of knowledge, that is, knowledge about a portion of the data production process, the significance of knowledge about data collection is clear. For each data quality dimension, two of the three data collection knowledge variables are significant. This finding has practical importance for managers because this is the least visible or important of the three domains in terms of managerial attention. In designing any information system, much of the importance is placed on requirements analysis (what users need in their utilization process) and on designing the computer system (what the data storage process is). Typically, little attention is given to the data collection process, which is generally the responsibility of clerks or is a by-product of organizational transactions [12].

4.3 Patterns by Knowledge Mode

All modes of knowledge, i.e., knowing-what, knowing-how, and knowing-why, are significant for producing high-quality data. Relevancy is the only dimension that does not have at least one significant knowledge variable in each mode. It does not have a significant knowing-how variable.

Beyond this, the clear pattern is that knowing-what and knowing-how are always positively associated with data quality, but knowing-why is always negatively associated with data quality. Those respondents who assessed their knowing-what and knowing-how as high, assessed data quality as high. On the other hand, those respondents who assessed their knowing-why high, assessed data quality as low.

How do we account for these different patterns of knowing-what and knowing-how vs. knowing-why? Specifically, the association of high knowing-why with low data quality needs an explanation. An explanation for this phenomenon may lie in the characteristics of the knowledge mode, knowing-why. By definition, knowing-why is a deeper form of understanding than knowing-what and knowing-how, in terms of understanding problems. Those with high knowing-why easily identify data quality problems and discover new ways of solving them. When assessing data quality, their assessment may be lower because they are aware of the problems with data. Thus, our conjecture is that the respondents with high knowing-why tend to assess data quality lower that those with high knowing-what and knowing-how. This explanation is somewhat supported by previous work that found that IT professionals assess the quality of data much higher than those who collect or use the data [7]. This might be attributable to lower why-knowledge of IT professionals about data collection and utilization processes. This conjecture of knowing-why as a differentiating mode is exploratory at this point and should be explored in a future study.

4.4 Variables other than Knowledge

Since it is possible that the role of the respondent, data collector, data custodian, or data consumer, might affect the results, we tested whether addition of role in the models was significant, but it was not. We also tested a company indicator to check whether results were significantly different by companies. For the timeliness dimension, the company indicator was significant indicating different levels of the dimension of timeliness in different companies.

5. CONCLUSION

In this paper, we used linear regression models to estimate the relationship between various modes and domains of knowledge about a data production process and the quality of the data produced by such a process. For the five data quality dimensions investigated, accuracy, relevancy, timeliness, completeness, and accessibility, different patterns of significant knowledge were important. The accuracy dimension had the most complex pattern of significant knowledge variables, while timeliness, completeness, and accessibility had similar patterns.

Overall, all three modes of knowledge, knowing-what, knowing-how, and knowing-why were significant. The interesting knowledge mode pattern was the positive relationships for knowing-what and knowing-how, and the negative relationships for knowing-why. While all three domains of knowledge, knowledge about data collection, about data storage, and about data utilization, were important, a key result is the overall importance of the knowledge about the data collection process.

These results have implications for managers who are seeking to improve data quality. Managers need to realize the complexity of knowledge for producing accurate data. Managers also should focus more on the data collection process. As the starting point of any data production process, it probably has more effect on the resulting quality of data then most managers realize. Furthermore, more knowing-why knowledge may contribute to lower assessments of the quality of data, which in turn provides a starting point for thinking about how to improve data quality.

For researchers, this study raises as many questions as it answers. Future research should investigate further and verify our explanations for two interesting findings from this study, the importance of knowledge of the data collection process to data quality and the negative contribution of why-knowledge to data quality. Future research should investigate the other dimensions of data quality to determine whether the patterns of relationships observed for the five dimensions included in this paper also hold for the other dimensions, and should test the five models estimated in this paper with another set of data.

This study has two limitations that future research should investigate. First, this research is an exploratory investigation using a simple linear additive model. While an appropriate starting point, a more complex model might provide better explanations. Second, this study uses measures of data quality assessed by the same individual that provided the knowledge measures. Future research studies should test whether these relationships hold with other measures of data quality.

Acknowledgement

We thank Neema Isack and Ning Liu, who performed the statistical analysis in this paper as a project for their statistical consulting class, under the direction of Professor Joseph Petruccelli.

REFERENCES

1. Ballou, D.P., Wang, R.Y., Pazer, H. and Tayi, G.K. Modeling Information Manufacturing Systems to Determine Information Product Quality. *Management Science*, *44* (4). 1998, pp. 462-484.
2. Hammer, M. *The Agenda: What Every Business Must Do to Dominate the Decade*. Ransom House, New York, NY, 2003.
3. Huang, K., Lee, Y. and Wang, R. *Quality Information and Knowledge*. Prentice Hall, Upper Saddle River: N.J., 1999.
4. Kahn, B.K., Strong, D.M. and Wang, R.Y. Information Quality Benchmarks: Product and Service Performance *Communications of the ACM*, 2002, pp. 184-192.

5. Lee, Y. and Strong, D. Beyond Knowing-what and Knowing-how: An Inquiry into Knowing -why about Data Processes and their Quality. *Journal of Management Information Systems*. Forthcoming.
6. Lee, Y.W., Why "Know Why" Knowledge is Useful for Solving Information Quality Problems. in *Americas Conference on Information Systems*, (Phoenix, AZ, 1996), The Association of Information Systems (AIS), pp. 200-202.
7. Lee, Y.W., Strong, D.M., Kahn, B.K. and Wang, R.Y. AIMQ: A Methodology for Information Quality Assessment. *Information & Management*, *Vol. 40* (2). 2002, pp. pp. 133-146.
8. March, J.G. Exploration and Exploitation in Organizational Learning. *Organization Science*, *2* (1). 1991, pp. 71-87.
9. Schultz, U. and Leidner, D. Studying Knowledge Management in Information Systems Research: Discourses and Theoretical Assumptions. *Management Information Systems Quarterly*, *Vol. 26* (No. 3). 2002, pp. 1-30.
10. Strong, D.M., Lee, Y.W. and Wang, R.Y. Data Quality in Context. *Communications of the ACM*, *40* (5). 1997, pp. 103-110.
11. Wang, R.Y. A Product Perspective on Total Data Quality Management. *Communications of the ACM*, *41* (2). 1998, pp. 58-65.
12. Wang, R.Y., Lee, Y.L., Pipino, L. and Strong, D.M. Manage Your Information as a Product. *Sloan Management Review*, *39* (4). 1998, pp. 95-105.
13. Wang, R.Y. and Strong, D.M. Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, *12* (4). 1996, pp. 5-34.

APPENDIX A: REGRESSION RESULTS

DQ Dimension		**Adjusted R^2**	**F stat**	**p**
Accuracy		0.15	7.99	<0.0001
Variable*	**Estimate**	**Std. Error**	**t stat**	**p**
Intercept	3.275	0.508	6.45	<0.0001
KWT-S	0.408	0.123	3.31	0.001
KHW-C	0.408	0.180	2.27	0.024
KHW-U	0.192	0.107	1.79	0.075
KWY-C	-0.546	0.180	-3.04	0.003
KWY-S	-0.240	0.132	-1.82	0.070

* KWT knowing-what, KHW knowing-how, KWY knowing-why
-C about collection, -S about storage, -U about utilization

DQ Dimension		**Adjusted R^2**	**F stat**	**p**
Relevancy		0.23	21.12	<0.0001
Variable*	**Estimate**	**Std. Error**	**t stat**	**p**
Intercept	2.892	0.501	5.78	<0.0001
KWT-C	0.413	0.125	3.30	0.001

KWT-U	0.280	0.094	2.98	0.003
KWY-C	-0.198	0.112	-1.77	0.078

* KWT knowing-what, KHW knowing-how, KWY knowing-why
-C about collection, -S about storage, -U about utilization

DQ Dimension		**Adjusted R^2**	**F stat**	**p**
Timeliness		0.13	8.37	<0.0001
Variable*	**Estimate**	**Std. Error**	**t stat**	**p**
Intercept	4.378	0.945	4.63	<0.0001
KWT-S	0.244	0.083	2.95	0.004
KHW-C	0.477	0.169	2.83	0.005
KWY-C	-0.469	0.174	-2.69	0.008
Company	-0.182	0.109	-1.67	0.097

* KWT knowing-what, KHW knowing-how, KWY knowing-why
-C about collection, -S about storage, -U about utilization

DQ Dimension		**Adjusted R^2**	**F stat**	**p**
Completeness		0.17	14.77	<0.0001
Variable*	**Estimate**	**Std. Error**	**t stat**	**p**
Intercept	2.877	0.444	6.48	<0.0001
KWT-S	0.266	0.073	3.66	0.0003
KHW-C	0.496	0.149	3.33	0.001
KWY-C	-0.475	0.154	-3.09	0.002

* KWT knowing-what, KHW knowing-how, KWY knowing-why
-C about collection, -S about storage, -U about utilization

DQ Dimension		**Adjusted R^2**	**F stat**	**p**
Accessibility		0.16	13.50	<0.0001
Variable*	**Estimate**	**Std. Error**	**t stat**	**p**
Intercept	2.406	0.520	4.63	<0.0001
KWT-S	0.214	0.085	2.51	0.013
KHW-C	0.596	0.175	3.42	0.001
KWY-C	-0.451	0.180	-2.51	0.013

* KWT knowing-what, KHW knowing-how, KWY knowing-why
-C about collection, -S about storage, -U about utilization

APPENDIX B: MEASURES

Knowledge (Independent) Measures

All items are measured on a 1 to 10 scale where 1 is "Very small extent", 5 is "Average", and 10 is "Very Large Extent." Items labels with "(R)" are reverse coded.

Knowing what about data collection (5 items, Cronbach's Alpha=.92)
(KWTC01) I know who creates this information.
(KWTC02) I know which group collects this information.

(KWTC03) I know the procedures by which this information is collected.
(KWTC04) I know the steps taken to gather this information.
(KWTC06) I know the sources of this information.

Knowing what about data storage (6 items, Cronbach's Alpha=.94)
(KWTS01) I know who maintains this information in our computers.
(KWTS02) I know which group maintains this information in our computers.
(KWTS03) I know the procedures used to store this information in our computers.
(KWTS04) I know the steps taken to store and maintain this information in our computers.
(KWTS05) I know which of our computers stores this information.
(KWTS06) I know which software is used for storing this information in our computers.

Knowing what about data utilization (5 items, Cronbach's Alpha=.93)
(KWTU01) I know who (individual or group) uses this information.
(KWTU02) I know which group uses this information.
(KWTU03) I know the procedures in which this information is used.
(KWTU04) I know the steps taken when using this information.
(KWTU05) I know the tasks which require the use of this information.

Knowing how-to about data collection (5 items, Cronbach's Alpha=.94)
(KHWC01) When typical problems arise with collecting this information, I know how we handle them.
(KHWC02) I know the usual solutions for problems with collecting this information.
(KHWC03) I know how to fix routine problems with collecting this information.
(KHWC04) I know how to fix recurring problems with collecting this information.
(KHWC05) I know the standard procedures for correcting deficiencies in information when collecting it.

Knowing how-to about data storage (5 items, Cronbach's Alpha=.96)
(KHWS01) When typical problems arise with storing this information in our computers, I know how we handle them.
(KHWS02) I know the usual solutions for problems with storing this information in our computers.
(KHWS03) I know how to fix routine problems with storing this information in our computers.
(KHWS04) I know how to fix recurring problems with storing this information in our computers.
(KHWS05) I know our standard procedures for correcting deficiencies in information when storing it in our computers.

Knowing how-to about data utilization (4 items, Cronbach's Alpha=.94)
(KHWU01) When typical problems, such as interpretation or access, arise with using this information, I know how we handle them.
(KHWU02) I know the usual solutions for problems with using this information.
(KHWU03) I know how to fix routine problems with using this information.
(KHWU05) I know our standard procedures for correcting deficiencies in information when using it.

Knowing why about data collection (4 items, Cronbach's Alpha=.91)
(KWYC02) I know the problems encountered in collecting this information.
(KWYC04) I understand the information collection procedures well enough to recognize why this information is collected incorrectly.
(KWYC05) I can detect sources of new problems in collecting this information.
(KWYC06) I can recognize new problems as they arise in collecting this information.

Knowing why about data storage (7 items, Cronbach's Alpha=.93)
(KWYS02) I know why this information is displayed in this form in our computers.
(KWYS03) I know some of the problems in storing this information appropriately in our computers.
(KWYS04) I know why it is difficult to store this information in our computers in an easy-to-interpret manner.
(KWYS06) I understand our computing environment well enough to analyze why this information is stored inadequately.
(KWYS08) I can recognize new problems as they arise in storing and maintaining this information in our computers.
(KWYS10) I know why people have difficulty with computer access procedures for this information.
(KWYS11) I know why it is difficult to store all this information in our computers.

Knowing why about data utilization (6 items, Cronbach's Alpha=.88)
(KWYU03) I know some of the problems in ensuring that this information is used appropriately.
(KWYU06) I can detect sources of new problems in using this information.
(KWYU07) I can recognize new problems as they arise in using this information in a new task.
(KWYU15) I cannot diagnose problems in using this information. (R)
(KWYU16) I cannot find the causes of new problems in the use of this information. (R)
(KWYU17) I cannot recognize when new problems arise in using this information in a new task. (R)

Data Quality (Dependent) Measures

All items are measured on a 0 to 10 scale where 0 is "Not at all", 5 is "Average", and 10 is "Completely." Items labels with "(R)" are reverse coded.

Accuracy. (4 items, Cronbach's Alpha=.91)
This information is correct.
This information is incorrect. (R)
This information is accurate.
This information is reliable.

Completeness. (6 items, Cronbach's Alpha=.87)
This information includes all necessary values.
This information is incomplete. (R)
This information is complete.
This information is sufficiently complete for our needs.
This information covers the needs of our tasks.
This information has sufficient breadth and depth for our tasks.

Timeliness. (5 items, Cronbach's Alpha=.88)
This information is sufficiently current for our work.
This information is not sufficiently timely. (R)
This information is not sufficiently current for our work. (R)
This information is sufficiently timely.
This information is sufficiently up-to-date for our work.

Relevancy. (4 items, Cronbach's Alpha=.94)
This information is useful to our work.
This information is relevant to our work.
This information is appropriate for our work.
This information is applicable to our work.

Accessibility. (4 items, Cronbach's Alpha=.92)
This information is easily retrievable.
This information is easily accessible.
This information is easily obtainable.
This information is quickly accessible when needed.